DOCUMENT RESUME

ED 241 549 TM 840 021

AUTHOR Gamache, LeAnn M.; Novick, Melvin R.
TITLE Choice of Variables and Gender Differentiated Prediction within Selected Academic Programs. Research Report #105.
INSTITUTION Iowa Univ., Iowa City. Evaluation and Examination Service.
PUB DATE 10 Nov 83
NOTE 34p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *College Entrance Examinations; Grade Point Average; Higher Education; *Mathematical Models; *Predictive Measurement; *Predictor Variables; Scores; *Sex Differences; Sex Discrimination; Test Bias
IDENTIFIERS American College Testing Program; *Johnson Neyman Technique

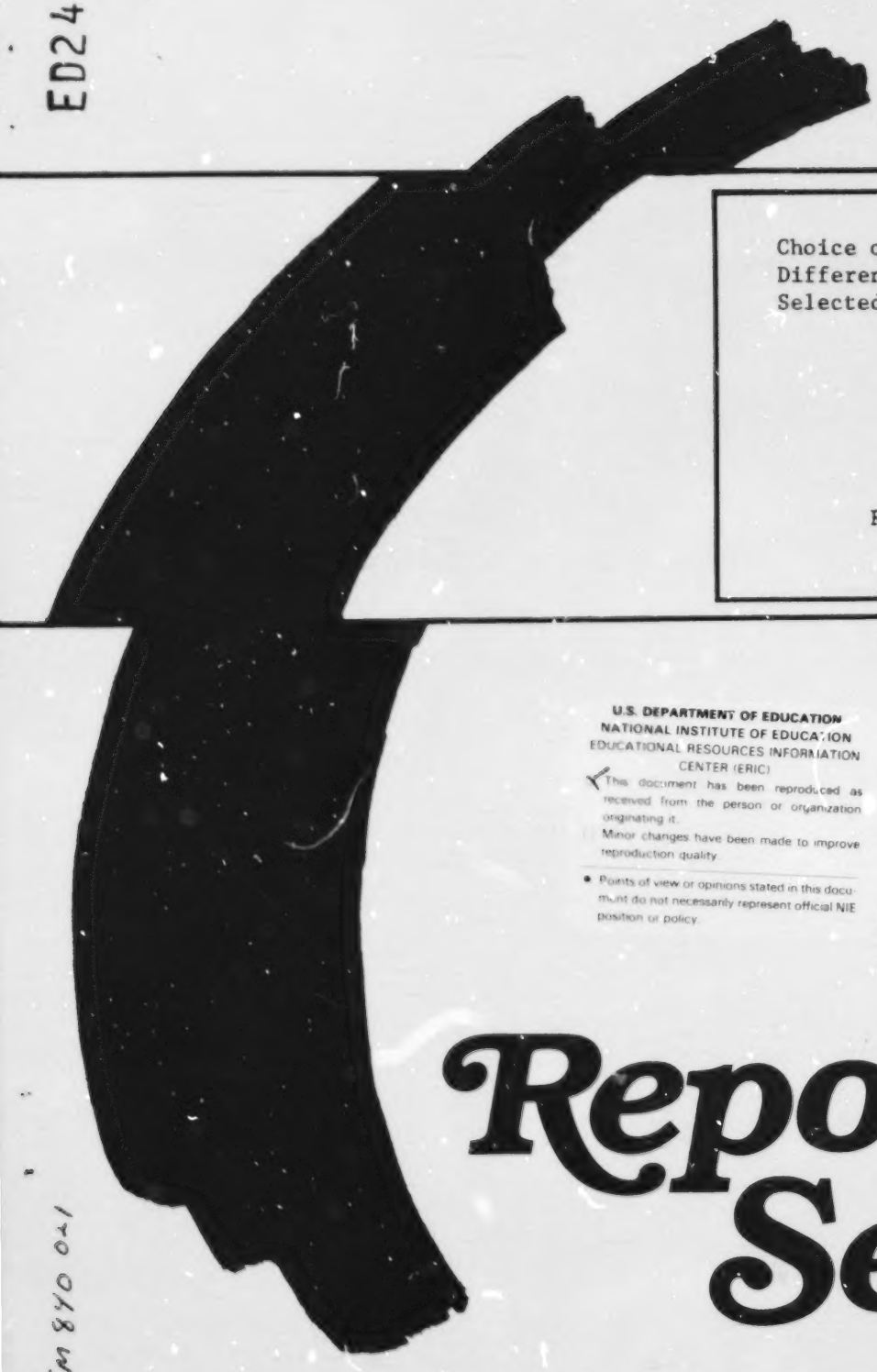
ABSTRACT

The existence of differential prediction of two-year grade point average is reported for gender groups within programs of study at the University of Iowa. Academic records of all freshmen entering the University in 1978 in the fields of Business, Liberal Arts, Pre-Medicine, and those undecided as to major were analyzed with respect to American College Testing (ACT) Program Composite score and the ACT Mathematics Usage, English Usage, Social Studies Reading, and Natural Science Reading subtest scores, using a two-year cumulative grade point average as the criterion of academic success. Johnson-Neyman analyses indicate regions on the predictor score scales where differential prediction has practical impact within each program. A large portion of students is generally affected. A methodology is outlined for finding a prediction algorithm to reduce differential prediction while maintaining prediction potency. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED241549

Evaluation & Examination Service
The University of Iowa



Choice of Variables and Gender Differentiated Prediction Within Selected Academic Programs

LeAnn M. Gamache
and
Melvin R. Novick

November 10, 1983

Research Report #105

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

L. M. Gamache

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Report Series

TM 840 041

Abstract

The existence of differential prediction of two-year grade point average is reported for gender groups within programs of study at a large state university. Johnson-Neyman analyses indicate regions on the predictor score scales where differential prediction has practical impact within each program. A large portion of students is generally affected. A methodology is outlined for finding a prediction algorithm to reduce differential prediction while maintaining prediction potency.

INTRODUCTION

The use of ability tests in the selection process has increasingly come under attack in the last two decades from test takers, consumer advocate groups, and from government agencies such as the EEOC (which is charged with enforcement of Title VII regulations to eliminate employment discrimination). Most criticism is made with respect to issues of test quality, conflict among the goals of the test developer, test user, and test taker, and issues of selection, bias. In addition, concern regarding adverse impact upon "protected" groups is found in the employment arena where selection test results can directly affect one's immediate source of livelihood. The issues of adverse impact, fairness, and bias in selection procedures within the educational system are also seen as important, particularly by those who view education as the primary means for upward mobility, and educational testing, in turn, as the gatekeeper for educational opportunity.

This paper does not deal with the issues of adverse impact or fairness in testing since these are nontechnical, policy issues; instead the narrower, technical issue of bias in testing is addressed. By definition, no predictive bias exists if the respective predictive relationships for defined minority, gender, or other groups of interest are given by a common algorithm such as a regression line and associated standard error of prediction (Cleary, 1968). If the regression slopes, intercepts, or errors of estimate differ for the groups but the same score interpretation is made without regard to group membership, the selection decision will be considered biased. Differing standard errors of estimate indicate that predictions of both the criterion performance and of the cut scores set on the predictor are made with differing degrees of confidence for the two groups. Differing regression slopes indicate that the test differentially predicts the criterion for the groups, depending on the designated critical level or cut score. Differing intercepts, given identical slopes, indicate the group with the larger intercept is systematically underpredicted by use of the common prediction equation. Substantial differences in the cut score needed to predict a specified level of the criterion can exist with even minor differences in regression slopes or intercepts, particularly if selection is made in one of the tails of the predictor distribution.

This paper addresses the issue of the existence of gender bias in prediction by examining the use of a major test at a large state university. By introducing certain controls (blocking on academic program) and certain

methodologies (Johnson-Neyman intervals) not generally used in this context, a more definitive answer is given that may help to generalize the available results on this topic. The issue of whether differential prediction can be reduced by judicious choice of predictor variables is also addressed. Future research should verify across institutions the results from this research found across programs of study.

EARLIER RESEARCH

It appears that well constructed cognitive ability tests appropriately selected and used for college admission generally do not exhibit predictive bias against blacks under the definition generally adopted (Linn, 1982). In studies of academic prediction, Linn (1982) indicates that when different prediction equations do exist, the error is generally one of overprediction for black students (Breland, 1978; Cleary, 1968; Davis & Kerner-Hoeg, 1971; Temp, 1971).

Research in educational contexts concerning the differential prediction of such tests for other minority groups is more limited (see Duran, 1983; Goldman and Hewitt, 1976). With respect to differential prediction for gender groups, Linn (1982) reports that the standard error of estimate is generally smaller for women when traditional predictors are used to predict freshman GPA (ACT, 1973). He further indicates that systematic errors of underprediction of women have been consistently found with the use of a common prediction equation for gender groups. In results summarized by Linn (1973), GPA scores for women having SAT scores at the female average were underpredicted by one-fourth to a full standard deviation below their actual mean GPA. The equation for men underpredicted the women's GPA by .36 points. Similar results were obtained from prediction of GPA from ACT scores. Underprediction of women with ACT scores at the mean averaged .27 grade points over 19 institutions. When high school grades were combined with ACT scores, differential prediction was reduced slightly to .20 points (ACT, 1973). In military personnel testing gender bias is also found. The 1983 Biennial Report of the Defense Advisory Committee on Military Personnel Testing (Linn, 1983, p.4) indicates that "Differential prediction results on the basis of gender show consistent and rather substantial differences . . . that apparently result in underprediction of the criterion performance of women."

The amount of systematic error in prediction from a common equation can be large relative to differences in GPA given different predictor scores. The effect of the differential prediction in GPA from SAT for women is approximately as large as that predicted between women with

average SAT scores and those with an SAT score one standard deviation below the mean (Linn, 1982).

Using essay performance as the criterion measure, Breland and Griswold (1982) examined differential prediction from the SAT-Verbal, SAT-Math, the Test of Standard Written English, and three subtests of the English Placement Test for gender groups and for four ethnic groups. Results were generally consistent with the findings of previous research. Female performance on the criterion tended to be underestimated and minority groups' performance overestimated by these measures. A major contribution of the study was the attention to and attempt to control possible sources of criterion bias. Use of a performance test as the criterion made possible external scoring and a more consistent standard for assigning scores, thus eliminating possible bias in the criterion due to differential coursework. However, a problem exists with use of the essay criterion. Although writing skill may be an important aspect of achievement, it can at best be only a partial indicator of academic success. The need to control for field of study is further indicated by the finding of differential grading standards across fields (Goldman and Hewitt, 1975).

The need for additional research on the question of predictive bias of educational selection tests stems not only from the requirements of a specific institution using a specific test but from the fact that earlier studies did not control the confounding effects of different curricula for gender groups, did not make precise the regions in which reliable differences in prediction could be expected, and did not seek an alternate prediction algorithm that might reduce or eliminate differential prediction.

The purposes of this study are to (1) examine the existence of differential prediction of GPA from ACT subtest and composite scores for gender groups within four major programs at a large state university to control, in part, for differential coursework, (2) determine regions on the prediction variable or composite in which differential prediction occurs, and (3) determine whether differential prediction can be reduced through careful selection of a subset of ACT subtest scores to form a common algorithm. The issue of using weights other than regression weights to reduce differential prediction is not addressed in this paper. The expectation is that the methodology will apply beyond the institution in which the study was conducted.

METHODOLOGY

Academic records of all undergraduate freshmen entering the University in 1978 in the fields of Business (BUS), Liberal Arts (LIB), Pre-Medicine (MED), and those undecided as to major (UND) were analyzed with respect to the ACT Composite score (COMP) and the ACT Mathematics usage (MATH), English usage (ENG), Social Studies Reading (SS), and Natural Science Reading (NS) subtest scores, using a two-year cumulative grade point average (GPA) as the criterion of academic success. ACT scores are theoretically available for instructor review but were assumed to be rarely consulted. Thus this form of criterion contamination should not be present. The effect of differential coursework was partially controlled because analyses were made within each college. No administrative selection process exists for these programs. Self-selection into a program, however, was surely present and directly related to previous achievement but not taken into account in this analysis.

Hypotheses of coincidence of gender specific regression lines were investigated using classical hypothesis testing procedures with COMP, ENG, MATH, the optimally weighted set of all four ACT subtest scores, and the optimally weighted combination of ENG and MATH. Testing for coincidence of regression lines simultaneously tested hypotheses of equal slope and equal intercept. Testing the hypothesis of equal slope served as a follow-up in search of the source of any predictive bias. An alpha level of .10 was used because Type II error rate is more critical than Type I in the study of differential prediction.

Comparison of GPA predicted from gender specific equations at specific predictor values was made for selected predictor variables to determine the magnitude and direction of differential prediction within each program. Finally, Johnson-Neyman analyses (Pedhazur, 1982) were conducted to determine the regions of rejection of the null hypothesis of no difference along the predictor score scales. Such analyses indicate where group differences would be reliably found across replications. The proportion of the total potential applicant pool (the entering freshman class) and the proportion of women statistically affected by the differential prediction were determined using z-score transformations of boundary scores based on means and standard deviations for the freshman class. The ENG and MATH variables were hypothesized to be the subset of predictors best minimizing the differential prediction of GPA while maintaining the potency of the prediction.

RESULTS

Table 1 provides the sample sizes, means, and standard deviations by gender group within college for the ACT composite, the subtest scores, and GPA. In all programs the mean ENG score was higher for females. All other ACT score means were higher for males. Mean GPA was higher for females in all programs except Pre-Medicine.

Insert Table 1 about here

The contribution to the prediction of GPA by ACT score, gender, and the interaction of gender with ACT score was examined by program for the COMP, ENG, and MATH scores. The full prediction models determined for each of these predictors are given in Table 2 with the respective standard errors of estimate and squared correlations indicating the potency of the prediction. Gender specific equations can be obtained by substituting 0 and 1 for males and females respectively, then combining like terms (Kleinbaum & Kupper, 1978). As is shown in Table 2, the gender variable most affects prediction of GPA in the programs of Business and Pre-Medicine when the ENG subtest score is used as the sole predictor variable and in the Pre-Medicine program when the MATH score is used as the sole predictor. The regression equations for the combined groups are given in Table 3. Prediction for the Undecided category is relatively poor in that at most only 11% of the variance in GPA can be accounted for by any one of the single predictors.

Insert Tables 2 and 3 about here

Predictive models of cumulative GPA were fit within each program from the set of four ACT subtest scores, gender, and the gender-subtest score interactions. Raw score weights for these variables are given in Table 4a with the standard errors of estimate, squared multiple correlations, and gender specific equations within program. A particularly large gender weight is found for this set of predictors within Business. In that program, gender has an impact of .47 grade points when all subtest scores are used as predictors. Table 4b provides the regression equations for the combined groups. Note that some negative weights are found for Social Studies Reading (SS) and Natural Science Reading (NS) in both tables. The stability of such weights is questionable. Some of the negative weights may be due to sampling error, others due to self-selection. In

any event, their existence argues for not using these variables in the prediction.

Insert Table 4 about here

Simultaneous testing of the hypothesis of coincidence of regression lines for gender groups is equivalent to testing the significance of the contribution of gender and its interaction terms to the prediction of GPA given the ACT predictor(s). Testing the point hypothesis of parallel regression lines is equivalent to a test of the contribution of the gender-predictor(s) interactions (Kleinbaum & Kupper, 1978). Results of testing these hypotheses for each prediction model are given by program in Table 5. Differences expected to be reliable across replications were found for eight of the twelve single predictor models and for three of the four multiple predictor models. Table 5 indicates that in 11 of 16 comparisons the null hypotheses of coincidence of regression for males and females was rejected. It is particularly interesting that noncoincidence is present for the Math variable in all four programs. Following rejection of the null hypotheses of coincidence of the gender specific regression, a rejection of the null hypothesis of no difference in slope for gender specific equations was found only in the Pre-Medicine program when all subtest scores were used in the prediction equation.

Insert Table 5 about here

The direction and magnitude of the difference in prediction is best seen from the difference in predicted GPA determined from the separate regression equations. Predictions of GPA are thus reported in Tables 6a through 6d using the gender specific equations at scores closest to one standard deviation below the mean, the mean, and one standard deviation above the mean of the potential applicant pool, by program, for each of the single predictors and for the multiple subtest predictors. As a relative index of the magnitude and direction of the difference in prediction, the ratio of the difference between the predicted scores to the standard deviation for women within the respective program is provided.

Insert Table 6 about here

Tables 6a through 6d indicate that the differences in the regressions are of sufficient magnitude to be of practical importance. For example, in the Business program the difference in the predicted GPA using MATH at one standard deviation above the pooled mean is .19 grade point units greater for women than men using the gender differentiated regressions. In almost all instances use of the gender differentiated predictions increases the predicted criterion value for women. The greatest effect within the three programs for declared majors is for women at one standard deviation below the mean using the Math variable as the sole predictor. The difference is .30 grade point units. It might be argued that this is unimportant since few students at this level would be considered for selection. Recall, however, that explicit selection is not great for the programs. For those undecided as to major, the magnitude of the differential prediction is even greater. Women at one standard deviation above the pooled mean are underpredicted by .34 grade point units using MATH as the sole predictor and by .36 units when COMP is used as the sole predictor. Such differences represent approximately one-half the standard deviation of women in that group.

In several cases the obtained gender differentiated predictions are higher for men than women. The effects are small except at the lower levels of the English variable for Pre-Medicine. Note, however, that in this case the null hypothesis of coincidence was not rejected.

The largest differences occur among the undecided group. This weakly supports the hypothesis that some of the observed differences in prediction are due to differences in coursework. However, analysis from the previous tables suggests that even when coursework is more uniform, gender differences of substantial magnitude remain.

Insert Table 7 about here

Results of the Johnson-Neyman analysis using a difference of zero and an alpha level of .10, as given in Table 7, identify regions on each predictor scale in which the null hypothesis of coincidence is rejected. Only the use of ENG as the sole predictor within the Business and

Liberal Arts programs resulted in acceptance of the null hypothesis over the entire region of possible predictor scores. This predictor is the least potent in terms of accounting for the variance in GPA, however. With the use of COMP and of ENG within Pre-Medicine, the Johnson-Neyman procedure rejected the null hypothesis for some region of the predictor scale, while the null hypothesis was retained under one of the overall tests. It should be noted that the regions of rejection in each case involve sets of values that are relevant to student decision making. This result demonstrates the greater power of the Johnson-Neyman technique in comparison to the more standard hypothesis tests.

The z-score equivalents of boundary scores for the regions were determined and the proportions of the total freshman group and of the female subgroup, respectively, within these regions were found. These proportions are given as the proportion of the potential applicant pool (all entering freshmen) and of women in that pool expected to be affected by the differential prediction. The direction of the effect is not indicated by the Johnson-Neyman analysis, but Tables 6a through 6d generally indicate underprediction of criterion performance for women.

In the attempt to find a set of predictors resulting in less differential prediction, the analyses were repeated using a combination of ENG and MATH as predictors of CPA. Results of these analyses are provided in Tables 8 and 9. Table 8 reports the coefficients for the ACT subtests, gender, and gender-subtest score interaction variables for the full predictive model by college, as well as the regression equations, standard errors of estimate, and squared multiple correlations for the gender groups and the combined group. As would be expected, in most instances the standard error is less than that using any of the single predictors but generally slightly greater than using all four subtests as predictors. Comparison with Table 4 indicates only a small reduction (slightly more than .01 on the average) in the squared multiple correlations. This reduction is of little practical importance. Tests of hypotheses of differential regression, Table 9, indicate some improvement over both the earlier use of single predictors and of the set of all subtests as predictors. No evidence of differential prediction is found for the Liberal Arts or Pre-Medicine programs using this set of predictors.

Insert Tables 8 through 10 about here

A more complete and informative indication of improvement, however, requires a comparison of GPA predicted from the gender differentiated equations to determine magnitude and direction of bias. Table 10a indicates the GPA in hundreths predicted from the gender specific equations using the combination of ENG and MATH as predictors. Comparison with the differences in predictions reported in Tables 6a through 6d reveals substantial reduction in the magnitude of the differential prediction in the programs of Business and Liberal Arts over the sole use of COMP, of MATH, or the combined use of all subtest scores. However, females remain consistently underpredicted in all programs. If it is assumed that a difference of .10 grade point units is meaningful, an important underprediction of women still exists within each program at some level of the predictor scores for the potential applicant pool. Use of the two predictor equation reduced the degree of differential prediction; it did not solve the problem.

DISCUSSION

Use of a Common Algorithm

A practitioner may regard some of these findings as irrelevant, arguing that differential prediction would not be as evident in practice since equations determined from the combined group are used to predict criterion performance rather than predicting female performance with the equation based solely on males. There are at least two considerations to be made in evaluating this argument. First, although the combined group regression line will lie between the gender specific lines, its location is a function of the proportion of males (or females) in the combined group. The higher the proportion of males, the closer the combined group regression line will be to the regression line for males. Because the combined group equation will always be between the group specific equations, use of the combined group equation obscures the magnitude of the true difference because the fit criterion gives predominate weight to the more numerous subpopulation. This may result in limiting the statistical evidence of differential prediction to the extent that the gender groups are differentially represented in the combined group. Thus, in situations where selection procedures (self- or otherwise) have led to disproportionate representation of a group and differential prediction of the criterion performance exists, use of the

combined group regression equation may only perpetuate the system of selection bias by causing underselection of the underpredicted and underrepresented group. Such situations are not uncommon in academic and military selection.

The second problem with the argument is perhaps a philosophical one. Although use of the combined group equation in place of the male group equation usually reduces the evidence of differential prediction, the problem of systematic over- or underprediction of one group may still remain. If accuracy of prediction is the goal, as is the case when selection is made in order of predicted score, it seems most reasonable to reduce error in the prediction system overall. This is accomplished through the use of separate equations in situations of differential prediction. However, the use of a combined equation and common cut score has considerable appeal. It may be neither administratively feasible nor defensible to do otherwise, since use of different equations or cut scores appears to give advantage to one group.

For purposes of illustration, the values of the criterion predicted at three score levels from the combined group equation and the associated differences in predicted values are provided in Table 6, along with those discussed earlier using the gender specific equations. Although the amount of the differential prediction is reduced in most settings, a meaningful difference (if defined as .10 grade point units) still exists in many of the situations. The differential is reduced somewhat with the use of the regression weighted composite and the unweighted composite of ENG and MATH as shown in Tables 10a and 10b.

The Search For An Acceptable Algorithm

One approach to finding an algorithm to reduce differential prediction would involve a series of steps leading to a decision regarding whether a combined equation or separate equations should be used, and what set of variables and weights would be most appropriate. Following selection of appropriate predictors, the first step would be examination of differential prediction using the group specific equations and an alpha level set to control Type II error. Johnson-Neyman regions would be determined to identify differential prediction around potential cut scores. If no differential prediction is identified, a common equation would be used.

If differential prediction was evidenced, each predictor should be examined with respect to its contribution to differential prediction and to the potency of the prediction

equation. If a predictor variable could be found which contributed to the differential prediction, but made little unique contribution to the accuracy of prediction of criterion performance, that variable should be excluded and the examination returned to step one above.

If no variable could be eliminated without a meaningful loss in prediction accuracy, a choice must be made between minimizing error of prediction and using a common algorithm. If the goal is to maximize prediction accuracy, separate regression equations should be used based on the original set of variables. If the goal is to use a common equation, then those variables contributing most to differential prediction should be replaced by other less offending variables or weights. The combined group equation using the most potent of these less offending variables or weights should then be used. These steps are summarized by the decision tree given in Figure 1.

Insert Figure 1 about here

Table 11 reports results of the application of this procedure to the determination of the most appropriate algorithm for predicting GPA for entering freshmen using common weights across the four programs. The use of a single composite across the four programs seems particularly desirable because many entering freshmen are either undecided as to major or change their major before graduation. Regressions of GPA on COMP for gender groups pooled across programs were used to calculate predicted GPA at various score levels for the entering freshman class. Meaningful differential prediction existed at each of these score levels. Under use of the pooled equation, a meaningful level of differential prediction remained at the upper score levels. Values of the squared multiple correlation were .184, .215, and .19 respectively for the male equation, female equation, and pooled equation. Due to previous experience it was believed that SS and NS contributed to the differential prediction, while contributing relatively little to the potency of the prediction of two-year GPA, and thus could be omitted as predictors. This is not to say that these subtest scores are not of value. Indeed, they have been shown to be of considerable value in placement decision-making and in planning of individual courses of study.

The unweighted sum of the ENG and MATH variables was thus investigated for differential prediction. Values of

the squared multiple correlation were .207, .218, and .208 respectively for the male, female, and pooled equations. Again, differential prediction existed at a meaningful level at all score levels used. However, under use of the pooled algorithm no meaningful differential was reflected at these score levels. Thus the differential prediction was reduced by selecting a subset of predictors to enable use of a common algorithm without loss of prediction potency. Indeed a small increase in predictive power was observed.

Insert Table 11 about here

The reason for this reduction in differential prediction is that the use of ENG and MATH generally resulted in higher within gender predictions for males and somewhat lower predictions for females than the use of COMP. Overall, the benefit to women of using the two variable composite is negative at the lower levels and positive at the higher score levels. Thus in this study, an increase in overall prediction and a decrease in differential prediction was of marginal but positive benefit to women as a group, if the environment is considered to be competitive.

CONCLUSION

Findings from this study indicate that differential prediction within these programs, with a bias against women, exists when prediction of two-year cumulative GPA is made using the ACT Composite score, ACT Mathematics usage, or ACT English usage scores as single predictors or when using the complete set of ACT subtest scores, even when the effect of differential coursework is partially controlled. Johnson-Neyman regions indicate the practical impact of the differential prediction in these programs. A large proportion of students is generally affected.

The differential prediction was meaningfully reduced in two of the four programs with the use of a subset of the original four predictors. This reduction in differential prediction was accompanied by a negligible gain in predictive accuracy. However, differential prediction consistently underpredicting women remained and exceeded a minimal meaningful difference at some score level within each program. Depending on administrative considerations, separate regression equations for males and females within or across programs might be considered for use in prediction. Absent a consensus for such action, a practical alternative would be to use a weighted composite of the ENG

and MATH variables within each program or an unweighted composite of these variables for all students. In the present application, use of the unweighted composite may be the only practical recommendation due to the advantages of simplicity, apparent evenhandedness, a small increase in predictive power, and a small decrease in differential prediction.

Acknowledgment

We are indebted to Dr. T. Anne Cleary, Director of The University of Iowa Examination Service, for encouraging the undertaking of this research, supplying the data, and providing valuable commentary on successive drafts.

Table 1

Means and Standard Deviations of Criterion and Predictors by Program

	Males		Females	
Business	Mean	S.D.	Mean	S.D.
	N=321		N=202	
GPA	252.8	70.8	267.0	74.2
ENG	19.5	4.1	21.2	3.9
MATH	23.3	6.3	23.0	6.2
SS	21.8	5.6	20.4	6.2
NS	24.6	5.0	23.2	5.6
COMP	22.4	4.4	22.1	4.6
Liberal Arts				
	N=256		N=366	
GPA	257.4	78.6	260.5	82.8
ENG	21.3	4.9	21.9	4.1
MATH	23.4	7.0	20.4	6.7
SS	23.5	6.0	21.6	5.8
NS	26.2	5.6	23.9	5.3
COMP	23.7	5.0	22.1	4.6
Pre-Medicine				
	N=159		N=239	
GPA	278.0	73.0	273.1	73.0
ENG	21.1	3.8	21.9	4.4
MATH	26.5	4.9	23.5	6.2
SS	24.3	5.2	21.6	6.2
NS	28.3	4.8	25.3	5.7
COMP	25.2	3.9	23.2	4.9
Undecided				
	N=280		N=337	
GPA	245.5	82.6	265.5	70.6
ENG	20.4	4.6	21.6	4.3
MATH	23.0	6.1	20.6	6.4
SS	22.5	6.2	20.9	6.2
NS	25.8	5.5	23.5	5.4
COMP	23.0	4.6	21.8	4.6

GPA is on a scale of 000-400. ACT scores are on a scale of 00-36.

Table 2

Full Models* Using Single Score Predictors

Business			
X=COMP	Y'=7.878(X)+23.523(Z)-0.294(XZ)+76.119		
S.E.:	M=61.677, F=65.499	R SQR:	M=.243, F=.224
X=ENG	Y'=7.353(X)+38.856(Z)-1.731(XZ)+109.123		
S.E.:	M=64.325, F=71.026	R SQR:	M=.177, F=.087
X=MATH	Y'=5.439(X)+5.459(Z)+0.452(XZ)+126.183		
S.E.:	M=62.075, F=64.688	R SQR:	M=.234, F=.243
Liberal Arts			
X=COMP	Y'=7.774(X)+14.556(Z)+0.065(XZ)+72.849		
S.E.:	M=68.271, F=74.718	R SQR:	M=.248, F=.187
X=ENG	Y'=6.698(X)-17.271(Z)+0.734(XZ)+114.702		
S.E.:	M=71.697, F=77.058	R SQR:	M=.171, F=.135
X=MATH	Y'=5.050(X)+17.836(Z)+0.035(XZ)+139.148		
S.E.:	M=70.424, F=75.632	R SQR:	M=.200, F=.167
Pre-Medicine			
X=COMP	Y'=8.082(X)-3.901(Z)+0.662(XZ)+74.234		
S.E.:	M=66.185, F=59.62	R SQR:	M=.183, F=.337
X=ENG	Y'=6.539(X)-46.477(Z)+1.656(XZ)+140.162		
S.E.:	M=68.779, F=63.649	R SQR:	M=.118, F=.244
S=MATH	Y'=7.696(X)+54.091(Z)-1.524(XZ)+73.959		
S.E.:	M=62.859, F=62.503	R SQR:	M=.263, F=.271
Undecided			
X=COMP	Y'=4.839(X)-3.758(Z)+1.363(XZ)+134.046		
S.E.:	M=79.691, F=64.525	R SQR:	M=.073, F=.166
X=ENG	Y'=5.262(X)+4.889(Z)+0.411(XZ)+138.280		
S.E.:	M=79.048, F=66.284	R SQR:	M=.087, F=.120
X=MATH	Y'=3.666(X)+17.009(Z)+0.579(XZ)+161.112		
S.E.:	M=79.649, F=65.279	R SQR:	M=.074, F=.147

X=Predictor score
Z=Gender (0 if male, 1 if female)
XZ=Interaction of Gender and Predictor

*Gender specific equations may be determined by substituting 0 or 1 for Z as appropriate and combining like terms.

Table 3

Regression Equations with Standard Error of Estimates for Combined Groups using Single Predictors

Business

X=COMP Y'=7.686(X)+86.964
S.E. = 63.600, R SQR = .129

X=ENG Y'=6.792(X)+121.295
S.E. = 66.959, R SQR = .146

X=MATH Y'=5.581(X)+129.017
S.E. = 63.464, R SQR = .233

Liberal Arts

X=COMP Y'=7.532(X)+87.766
S.E. = 72.442, R SQR = .202

X=ENG Y'=7.057(X)+106.211
S.E. = 74.801, R SQR = .149

X=MATH Y'=4.783(X)+155.845
S.E. = 73.962, R SQR = .168

Pre-Medicine

X=COMP Y'=8.26(X)+76.834
S.E. = 62.459, R SQR = .269

X=ENG Y'=7.517(X)+112.925
S.E. = 65.881, R SQR = .187

X=MATH Y'=6.301(X)+119.376
S.E. = 63.016, R SQR = .256

Undecided

X=COMP Y'=5.214(X)+139.836
S.E. = 72.972, R SQR = .01

X=ENG Y'=5.669(X)+137.236
S.E. = 72.549, R SQR = .110

X=MATH Y'=3.553(X)+179.373
S.E. = 73.511, R SQR = .087

Table 4

(a) Regression Weights for Components of Full Predictive Models of GPA Using All Subtests as Predictors*

	BUS	LIB	MED	UND
ENG	3.362	2.219	2.671	4.177
MATH	3.764	2.594	6.943	2.707
SS	0.330	1.378	0.664	0.774
NS	0.585	1.483	-1.380	-2.461
GENDER	47.148	-0.854	4.746	-6.284
GENG	-4.886	1.249	-0.009	-2.085
GMATH	1.074	0.534	-4.140	0.141
GSS	1.252	-2.070	0.644	0.905
GNS	0.743	0.817	3.572	1.984
CONSTANT	77.884	78.114	60.466	144.100

*Variables preceded with G are interactions with gender. Gender specific equations are obtained by setting G to 0 or 1 for males and females respectively then combining like terms.

S.E.	Males	60.936	68.684	62.819	78.095
	Females	64.227	74.005	59.607	64.150
R SQR	Males	.268	.248	.278	.119
	Females	.265	.209	.345	.183

(b) Combined Group Equations Using All Subtest Scores as Predictors

Business
Y'=2.455(E)+4.128(M)+0.491(S)+0.626(N)+87.666
S.E. = 62.667, R SQR = .256

Liberal Arts
Y'=3.694(E)+2.764(M)-0.196(S)+1.767(N)+79.83
S.E. = 72.010, R SQR: .215

Pre-Medicine
Y'=3.058(E)+4.309(M)+1.136(S)+0.237(N)+70.547
S.E. = 61.471, R SQR: .298

Undecided
Y'=4.299(E)+2.405(M)+1.017(S)-1.921(N)+139.035
S.E. = 71.442, R SQR: .141

Table 5

Tests of Hypotheses of Differential Prediction

	Ho:Parallel F	Ho:Parallel RejLvl	Ho:Coincident F	Ho:Coincident RejLvl
Business				
COMP	0.0555	.8138	4.5010*	.0115
ENG	1.2907	.2564	0.7869	.4558
MATH	0.2467	.6197	4.0550*	.0179
MULT	1.5161	.1962	2.4935*	.0302
Liberal Arts				
COMP	0.0028	.9577	3.6197*	.0302
ENG	0.2912	.5897	0.1720	.8421
MATH	0.0016	.9680	4.5980*	.0104
MULT	0.4384	.7809	1.4824	.1935
Pre-Medicine				
COMP	0.1872	.6655	1.8902	.1524
ENG	0.9805	.3227	1.8422	.1598
MATH	1.5672	.2114	3.3518*	.0360
MULT	2.3639*	.0526	2.4748*	.0318
Undecided				
COMP	1.1710	.2796	11.1151*	.0000
ENG	0.0991	.7530	2.6690*	.0701
MATH	0.3803	.5377	12.6997*	.0000
MULT	1.0182	.3972	3.1971*	.0074

Table 6

(a) Predicted GPA in Hundreths Using Various Predictors at Three Score Levels for Business

	Combined	M	F	M-F	C-F
COMP					
-1	218	210	229	-19	-11
M	264	257	274	-17	-10
+1	310	305	320	-15	-10
ENG					
-1	237	234	244	-10	- 7
M	264	264	266	- 2	- 2
+1	291	293	289	+ 4	+ 2
MATH					
-1	218	213	226	-13	- 8
M	257	251	267	-16	-10
+1	296	289	308	-19	-12
MULT					
-1	216	212	228	-16	-12
M	261	257	272	-15	-11
+1	305	302	316	-14	-11

(b) Predicted GPA in Hundreths Using Various Predictors at Three Score Levels for Liberal Arts

	Combined	M	F	M-F	C-F
COMP					
-1	216	205	221	-16	- 5
M	261	252	268	-16	- 7
+1	306	298	315	-17	- 9
ENG					
-1	226	229	224	+ 5	+ 2
M	254	255	253	+ 2	+ 1
+1	283	282	283	- 1	0
MATH					
-1	232	220	238	-18	- 6
M	266	255	274	-19	- 8
+1	299	291	310	-19	-11
MULT					
-1	219	209	224	-15	- 5
M	261	252	268	-16	- 7
+1	303	294	312	-18	- 9

(c) Predicted GPA in Hundreths Using Various Predictors at Three Score Levels for Pre-Medicine

	Combined	M	F	M-F	C-F
COMP					
-1	217	212	219	- 7	- 2
M	267	260	271	-11	- 4
+1	316	309	324	-15	- 8
ENG					
-1	241	251	233	+18	+ 8
M	271	277	266	+11	+ 5
+1	301	304	299	+ 5	+ 2
MATH					
-1	220	197	227	-30	- 7
M	264	251	270	-19	- 6
+1	308	305	313	- 8	- 5
MULT					
-1	214	200	220	-20	6
M	265	256	269	-13	- 4
+1	315	313	318	- 5	- 3

(d) Predicted GPA in Hundreths Using Various Predictors at Three Score Levels for Undecided

	Combined	M	F	M-F	C-F
COMP					
-1	228	216	236	-20	- 8
M	260	245	273	-28	-13
+1	291	274	310	-36	-19
ENG					
-1	234	228	240	-12	- 6
M	256	249	262	-13	- 6
+1	279	270	285	-15	- 6
MATH					
-1	236	220	246	-26	-10
M	261	245	276	-31	-15
+1	286	271	305	-34	-19
MULT					
-1	228	222	236	-14	- 8
M	259	250	272	-22	-13
+1	290	278	306	-28	-16

Table 7

Johnson-Neyman Regions
for Single Predictors

	Rejection Region	Proportion Affected* Total	Women
Business			
X=COMP	13.96<x<28.03	.[illegible]5	.852
X=ENG	No region	.000	.000
X=MATH	x>16.95	.819	.770
Liberal Arts			
X=COMP	16.65<x<30.06	.847	.843
X=ENG	No region	.000	.000
X=MATH	11.59<x<33.22	.902	.900
Pre-Medicine			
X=COMP	23.12<x<28.87	.382	.358
X=ENG	13.80<x<21.34	.471	.437
X=MATH	x<22.77	.773	.819
Undecided			
X=COMP	x>15.51	.944	.932
X=ENG	17.59<x<26.80	.702	.732
X=MATH	x>9.04	.983	.974

*Proportion of potential applicants affected by differential prediction.

Table 8

(a) Regression Weights for Full Predictive Models Using ENG and MATH as Predictors*

	BUS	LIB	MED	UND
ENG	3.836	3.691	2.297	3.747
MATH	4.062	3.518	6.807	2.122
GENDER	40.694	-5.279	26.833	13.004
GENG	-3.253	0.593	2.287	-0.537
GMATH	1.625	0.186	-2.688	0.937
CONSTANT	83.295	96.368	49.101	120.310

* Variables preceded with G are interactions with gender. Gender specific equations are obtained by setting G to 0 or 1 for males and females respectively then combining like terms.

S.E.	Males	60.811	69.069	52.585	78.433
	Females	64.822	74.235	60.613	64.330
R SQR	Males	.267	.234	.265	.105
	Females	.244	.200	.317	.174

(b) Combined Group Equations Using ENG and MATH as Predictors

Business
Total Y'=3.033(E)+4.512(M)+92.606
S.E. = 62.669, R SQR = .253

Liberal Arts
Total Y'=4.308(E)+3.337(M)+93.691
S.E. = 72.212, R SQR = .208

Pre-Medicine
Total Y'=3.917(E)+4.792(M)+72.195
S.E. = 61.547, R SQR = .292

Undecided
Total Y'=2.054(E)+4.207(M)+123.429
S.E. = 71.721, R SQR = .132

Table 9

Tests of Hypotheses of Differential Prediction
Using the English and Math Subtests as Predictors

	Ho:Parallel		Ho:Coincident	
	F	RejLvl	F	RejLvl
Business	1.9521	.1430	2.5800	.0529*
Liberal Arts	0.1800	.8350	1.3116	.2696
Pre-Medicine	1.8218	.1631	1.6083	.1869
Undecided	0.3700	.6909	4.7585	.0027*

Table 10

(a) Predicted GPA in Hundreths Using Weighted ENG and MATH as Predictors at Three Score Levels

	Combined	M	F	M-F	C-F
Business					
-1	216	213	225	-12	- 9
M	260	257	267	-10	- 7
+1	304	301	309	- 8	- 5
Liberal Arts					
-1	220	215	223	- 8	- 3
M	261	255	266	-11	- 5
+1	302	294	309	-15	- 7
Pre-Medicine					
-1	215	197	220	-23	- 5
M	265	254	267	-13	- 2
+1	314	311	314	- 3	0
Undecided					
-1	226	218	237	-19	-11
M	263	248	271	-23	- 8
+1	301	278	305	-27	- 4

(b) Predicted GPA in Hundreths Using the Unweighted Sum of ENG and MATH as Predictor at Three Score Levels

	Combined	M	F	M-F	C-F
Business					
-1	225	222	232	-10	- 7
M	261	258	266	- 8	- 5
+1	297	293	301	- 8	- 4
Liberal Arts					
-1	229	223	232	- 9	- 3
M	262	255	267	-12	- 5
+1	295	287	302	-15	- 7
Pre-Medicine					
-1	225	216	228	-12	- 3
M	265	260	267	- 7	- 2
+1	305	305	306	- 1	- 1
Undecided					
-1	234	222	243	-21	- 9
M	260	247	271	-24	-11
+1	286	272	299	-27	-13

Table 11

GPA Predicted From COMP and From Unweighted Composite of ENG and MATH at Various Score Levels Using Within Gender and Pooled Group Regressions Combining Data From Four Programs

Z-Score Level	Equation	COMP	ENG + MATH
-1.5	Male	203	203
	Female	219*	216*
	Pooled	213	210
-1.0	Male	220	220
	Female	236*	233*
	Pooled	229	228
-0.5	Male	236	238
	Female	254*	251*
	Pooled	246	245
0.0	Male	253	255
	Female	271*	269*
	Pooled	262	262
+0.5	Male	270	273
	Female	289**	286*
	Pooled	279	280
+1.0	Male	286	290
	Female	306**	304*
	Pooled	296	297
+1.5	Male	303	307
	Female	324**	322*
	Pooled	312	314

* Group adversely affected by differential prediction at meaningful level defined as .10 grade point units considering M-F comparison.

** Differential prediction under use of pooled equation adversely affects group.

Figure 1

Finding a Good Prediction Equation While
Reducing the Effect of Differential Prediction

Step 1: Examine DP on the basis of JN analysis or otherwise using separate equations.

Evidence of DP -----no-----> Use the common algorithm
|
| yes
|
v

Step 2: Delete and/or replace some variables and/or alter some weights and reexamine for DP as compared to results using the best common algorithm.

Evidence of DP -----no-----> Use the common algorithm
|
| yes
|
v

Step 3: Choose common or gender differentiated algorithms on the following basis.

a) Choose the best common algorithm if the loss in predictive accuracy or degree of differential prediction is tolerably small given administrative requirements.

b) Choose the gender differentiated algorithm if the gain in predictive accuracy is compellingly large or if the degree of gender differentiation is intolerably large with the best common algorithm.

BIBLIOGRAPHY

American College Testing Program. Assessing students on the way to college. Technical Report, 1973, Vol 1.

Breland, H.M. Population validity and college entrance measures. College Board Research and Development Report RDR78-79, No. 2, 1978.

Breland, H.M., & Griwsold, P.A. Use of a performance test as a criterion in a differential validity study. Journal of Educational Psychology, 1982, 5, 713-721.

Cleary, T.A. Test bias: prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement. 1968, 5, 115-124.

Davis J.A. & Kerner-Hoeg, S. Validity of pre-admission indices for Blacks and Whites in six traditionally white public universities in North Carolina. ETS Project Report PR-71-15. Princeton, N.J., 1971.

Duran, R.P. Hispanics' Education and Background, College Entrance Examination Board, New York, 1983.

Goldman, R. D. & Hewitt, B. N. Adaptation-level as an explanation of differential standards in college grading. Journal of Educational Measurement. 1975, 12, 149-161.

Goldman, R. D. & Hewitt, B. N. Predicting the success of Black, Chicano, Oriental and White college students. Journal of Educational Measurement. 1976, 13, 107-117.

Kleinbaum, D.G., & Kupper, L.L. Applied Regression Analysis and Other Multivariate Methods, North Scituate, Massachusetts: Duxbury Press, 1978.

Linn, R.L. Fair test use in selection. Review of Educational Research, 1973, 43, 139-161.

Linn, R. Ability Testing: Individual Differences, Prediction, and Differential Prediction. In Ability Testing: Uses, Consequences and Controversies. Committee on Ability Testing, National Academy Press, 1982.

Linn, R.L. Biennial Report of the Defense Advisory Committee on Military Personnel Testing, Office of the Secretary of Defense, 1983.

Linn, R.L. & Werts, C.E. Considerations for studies of test bias. Journal of Educational Measurement 1971, 8,

1-4.

Lord, F.M., & Novick, M.R. Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley, 1968.

Pedhazer, E. J. Multiple Regression in Behavioral Research, 2nd Ed., New York: CBS College Publishing, 1982.

Schrader, W.B. The predictive validity of College Board admissions tests. In W.H. Angoff, Ed. College Board Admissions Test Program: A Technical Report on Research and Development Activities Related to the SAT and Achievement Tests. New York, 1971.

Temp, G. Test bias: Validity of the SAT for blacks and whites in thirteen integrated institutions. Journal of Educational Measurement. 1971, 245-251.